

# Development and application of a recombination-based library versus library high-throughput yeast two-hybrid (RLL-Y2H) screening system

Fang Yang<sup>1,2,†</sup>, Yingying Lei<sup>3,4,†</sup>, Meiling Zhou<sup>3,4</sup>, Qili Yao<sup>3,4</sup>, Yichao Han<sup>5</sup>, Xiang Wu<sup>3,4</sup>, Wanshun Zhong<sup>1,2</sup>, Chenghang Zhu<sup>3,4</sup>, Weize Xu<sup>3,4</sup>, Ran Tao<sup>3,4</sup>, Xi Chen<sup>3,4</sup>, Da Lin<sup>3,4</sup>, Khaista Rahman<sup>3,4</sup>, Rohit Tyagi<sup>3,4</sup>, Zeshan Habib<sup>3,4</sup>, Shaobo Xiao<sup>3,4</sup>, Dang Wang<sup>3,4</sup>, Yang Yu<sup>6</sup>, Huanchun Chen<sup>3,4</sup>, Zhenfang Fu<sup>3,4,7</sup> and Gang Cao<sup>3,4,5,8,9,\*</sup>

<sup>1</sup>National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China, <sup>2</sup>College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China, <sup>3</sup>State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China, <sup>4</sup>College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, China, <sup>5</sup>College of Informatics, Huazhong Agricultural University, Wuhan 430070, China, <sup>6</sup>Key Laboratory of RNA Biology, Institute of Biophysics, CAS, Beijing 100101, China, <sup>7</sup>Departments of Pathology, College of Veterinary Medicine, University of Georgia, Athens, GA 30602, USA, <sup>8</sup>Bio-Medical Center, Huazhong Agricultural University, Wuhan 430070, China and <sup>9</sup>Key Laboratory of Development of Veterinary Diagnostic Products, Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China

Received July 06, 2017; Revised October 13, 2017; Editorial Decision November 07, 2017; Accepted November 15, 2017

## ABSTRACT

Protein-protein interaction (PPI) network maintains proper function of all organisms. Simple high-throughput technologies are desperately needed to delineate the landscape of PPI networks. While recent state-of-the-art yeast two-hybrid (Y2H) systems improved screening efficiency, either individual colony isolation, library preparation arrays, gene barcoding or massive sequencing are still required. Here, we developed a recombination-based ‘library vs library’ Y2H system (RLL-Y2H), by which multi-library screening can be accomplished in a single pool without any individual treatment. This system is based on the phiC31 integrase-mediated integration between bait and prey plasmids. The integrated fragments were digested by *MmeI* and subjected to deep sequencing to decode the interaction matrix. We applied this system to decipher the trans-kingdom interactome between *Mycobacterium tuberculosis* and host cells and further identified Rv2427c interfering with the phagosome-lysosome fusion. This concept can also be applied to other systems to screen

protein-RNA and protein-DNA interactions and delineate signaling landscape in cells.

## INTRODUCTION

As a fundamental milestone in the advancement of science, the human genome project has deciphered almost the entire gene sequences in our genetic instruction set and greatly promoted the progress of the current biomedical research. As the final product of most genes, proteins interact with each other and compose complicated and elegant networks to maintain the proper biological activities in all organisms. While the next generation sequencing technologies further facilitate genomic sequencing, the lack of high-throughput proteomic approaches for decoding proteomic networks hampers the systematical understanding of the molecular proteomic machinery of life. Generally, two approaches have been applied to investigate these intriguing protein networks. The proteomics-based approaches usually take advantage of mass spectrometry technologies to decipher the protein complex precipitated by a specific protein. Although this type of approach is straight forward, it is expensive and very challenging to detect low abundant proteins. Also, it is hard to distinguish direct and indirect interactions by this method (1). Protein array is another

\*To whom correspondence should be addressed. Tel: +86 15327266790; Email: gcao@mail.hzau.edu.cn

†These authors contributed equally to this work as first authors.

proteomic approach to identify protein interactions (2). Although it can be high-throughput, this method requires expensive and laborious protein purification processes. The genomic approaches, such as yeast two-hybrid (Y2H) system and phage display system, utilize cDNA to encode protein library and screen the library by various reporter systems.

Y2H is a powerful tool, which has been extensively applied to decipher gene function and signaling pathways in most organisms. During the last two decades, a lot of progress has been achieved for Y2H system to reduce the cost and labor for large-scale library screening (3–6). However, for most Y2H systems, it still needs to pick up individual colony or prepare arrays for specific yeast clones. Recently, several high-throughput Y2H systems such as eY2H, Smart-pooling Y2H, Y2H-Seq, Stitch-Seq Y2H, BFG-Y2H have been developed to improve the library screening efficiency (2–4,7–10). While these systems substantially increase the Y2H screening efficiency, the time-consuming and costly individual treatment, such as plasmid isolation and PCR amplification of individual positive colonies, individual arrays for library preparation, or individual gene barcoding, are still required. For example, the recently reported BFG-Y2H system utilizes Cre recombinase to tag each bait and prey genes with a specific barcode and identifies all interacting proteins simultaneously by high-throughput sequencing (9). However, the barcode fusion mediated by Cre recombinase can be reversible, it may influence the efficiency of the system. Nevertheless, BFG-Y2H system can easily decode the interaction network via reading these chimeric protein-pair barcodes from pooled Y2H-positive colonies. Since barcoding each bait and prey is required before screening, it is still laborious, expensive, and even not applicable when a large number of proteins need to be tested for interactions. A very recent study developed a CrY2H system for all-by-all high-throughput PPI screening. Instead of fusing barcodes for bait and prey genes, this system could directly link the interacting bait and prey pairs after screening (11). All of the PCR products of fused bait and prey genes were then randomly sheared and subjected to sequencing. Although this system does not require individual gene barcoding, it needs massively deep sequencing for big PPI network, as only a small fraction (2.4%) of the CrY2H reads can be mapped to both prey and bait genes (11). Thus, a simple, high-throughput and cost-effective Y2H library vs library system without any individual treatment is highly desired to explore the intricate and intriguing protein interaction landscape.

The microbes parasitized in host cells equip with two PPI networks. One constitutes of the proteins from the microbes themselves, and the other includes proteins from the microbes and host cells. The later trans-kingdom protein interaction network plays an important role in perturbing and modulating the host protein network, so that the microbes can escape the immunity and colonize host cell. *Mycobacterium tuberculosis* (*M.tb*) is an intracellular pathogen causing the most devastating zoonotic disease in the world (12). In 2015, 10.4 million new TB cases have been reported and 1.8 million died from this disease worldwide, which demonstrated TB became the leading infectious dis-

ease, even ranking above HIV/AIDS. During the process of infection and parasitization, the membrane and secretory proteins of *M.tb* extensively and specifically interact with host proteins to enter the cells, inhibit cellular anti-pathogen proteins, reprogram the host gene expression, and finally colonize the cells. One striking phenomenon is that *M.tb* could arrest phagosome maturation and escape into the cytosol for survival (13,14). In addition, the cytosolic *M.tb* perturbs host autophagy pathway by their secretory proteins, such as ESAT6, PknG (15,16). Deciphering the interactome between *M.tb* and host is pivotal to understand the molecular pathological basis of tuberculosis and facilitate the identification of anti-*M.tb* drug targets.

The aim of this study is therefore to develop a simple, low-cost and high-throughput library vs library Y2H screening system without any individual treatment and apply it to delineate the global trans-kingdom protein interaction network between the *M.tb* proteins (all the secretory and membrane genes) and the host proteins (TB susceptibility gene, immunity and autophagy-related genes).

## MATERIALS AND METHODS

### Plasmid construction

The bacterial attachment site ATTB and phage attachment site ATTP were inserted behind the BamHI site of Clontech GAL4 Y2H vector pGBKT7 and pGADT7, respectively, by recombination. The PhiC31 expression cassette Pol III:PhiC31:Cyc1 terminator was inserted into pGBKT7-ATTB and pGADT7-ATTP vector backbones at the site of AvrII and NheI, respectively, to generate mBD and mAD vectors for the following library construction. All insertions were confirmed by Sanger sequencing. The expression efficiency of the modified vectors was tested by comparing the transformation efficiency and interacting strength to the conventional ones using the known positive interacting pair, murine p53 and SV40 T-antigen. The PCR primer synthesis and DNA sequencing were performed by Shanghai Sangon Biotech, China.

### Library construction, screening and data Analysis

See Supplemental Experimental Protocol for RLL-Y2H.

### GFPph-mCherry-LC3 stable cell line construction and autophagy assay

Raw264.7 cells were maintained in a humidified incubator at 37°C with 5% CO<sub>2</sub> and grow in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS), 100 U/ml penicillin, and 100 mg/ml streptomycin. The H11 site homolog arms were amplified from the genomic DNA of Raw264.7 cells by PCR. The eGFP gene was mutated to a pH super-sensitive version using NEB site-direct mutagenesis kit according to the manufacturer's protocol. The mutated eGFP was then fused with mCherry-LC3 and homology arm by overlapping PCR (The sequence was shown in Supplemental Figure S6). The gRNA (5'-CACCTTTAGTCATAGTGTTG-3') targeted to the H11 site in the Raw264.7 cell was designed and cloned in CRISPR/Cas system vector px335 according to the

protocol from <http://www.genome-engineering.org/crispr>. GFPph-mCherry-LC3 with H11 site homology arms and px335-gRNA plasmid were co-transfected into Raw264.7 cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. The cells with GFPph-mCherry-LC3 reporter were selected by FACS (BD, CA, USA). For autophagy assay, the aggregation and intensity of GFP/mCherry fluorescence signal were measured by fluorescence microscopy (Olympus, Japan) and analyzed using ImageJ, as previously described (17).

### Statistical analysis

A Two-way *t*-test was used to analyze the significant differences of all the experiments by GraphPad Prism (version 5.01) software (San Diego, CA, USA). Differences were considered statistically significant when  $P < 0.05$ .

### Supplemental experimental protocol for RLL-Y2H

**Bait and prey library construction.** A total of 542 *M.tb* (H37Rv) membrane and secretory protein ORFs (a kind gift from Dr Xiao in Huazhong Agricultural university, based on the previous studies and prediction by *M.tb* genome website: <http://tuberculist.epfl.chand> <http://www.tbdb.org>) were amplified and cloned into mBD vector by recombination. Basically, PCR products with homologous arms and linearized mBD vector were together transformed into yeast strain Y2Hgold, and plated on SD/-T selective plates. The growing colonies were then amplified to confirm that a PCR fragment was integrated into mBD vector properly via Sanger sequencing or plasmid digestion.

The positive colonies for 542 ORFs were then separately mated with Y187 strain harboring empty mAD vector and were plated onto SD/-LT and SD/-AHLT selective plates to test their self-activation. Those not showing any self-activation were then pooled and used as the bait library.

A total of 316 host protein ORFs were amplified from BALB/c mice lung and brain cDNA. The ORFs include Autophagy-related genes ([www.mblintl.com](http://www.mblintl.com)), SNARE genes (Cell signaling technology, [www.cellsignal.com](http://www.cellsignal.com)), TB susceptible genes, Jak/Stat signaling, IL 6 receptor family, NF- $\kappa$ B signaling, T cell receptor signaling, as well as TLR signaling pathway (18–21). These ORFs were cloned into the mAD vector individually, and the subsequent plasmids were transformed into yeast strain Y187. Mixtures of transformants of every 20 plasmids were mated with Y2Hgold harboring murine p53 to rule out those showing sticky interaction or self-activation. The rest were pooled and used as the prey library.

**Y2H library screen.** Mate the bait and prey libraries at 50 rpm in  $4 \times 50$  ml of 2xYPDA medium for 24 h.

Next day, collect the cells and resuspend in 15 ml 0.9% NaCl, and plate onto 50 SD/-AHLT selective plates and grow for ~7 days, meanwhile plating several dilutions on SD/-L, SD/-T and SD/-LT selective plates to calculate the mating efficiency.

Pool all positive colonies from SD/-AHLT to isolate the integrated plasmids using the Yeast Plasmid Kit (OMEGA, D3376-01) according to the user manual.

**High-throughput DNA library construction.** Amplify the recombinated bait and prey fragments using the isolated integrated plasmids from the pooled positive colonies and amplify for 18–20 cycles. A smear band should be visible on agarose gel.

Digest the purified PCR products with *MmeI* (NEB, CA) which yields a ~110 bp target band, followed by low-melting gel purification.

Prepare Illumina P5 and P7 adapters by annealing overhang NN-nucleotides P5: 5'-ACACTCTTCCCTACACGACGCTCTTCCGATCTNN-3' and 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'; P7: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACT-3' and 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3', respectively.

Add P5 and P7 adapters to purified 110bp fragments by T4 ligation at 16°C for more than 2 h.

Amplify the ligation products using primers (5'-AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCT-3' and 5'-CAAGCAGAAGACGGCATACGAGAT(index)GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC-3') to construct high-throughput sequencing library.

The deep sequencing was performed on HiSeq2000 sequencer (Illumina, CA, USA).

**Data processing.** Scan all the reads for ATTL site 'TAGC GTGCGGGTGCCAGGGCGTGCCCTTGAGT TCTCTCAGTTGGGGGCGTTGAC' and *MmeI* recognition sites 'TCCGAC'.

Retain the reads containing about 20bp DNA flanking both sides of *MmeI*-ATTL-*MmeI* structure for subsequent analysis.

Trimmed the sequences to 13 bp from the ends of raw reads. Built the reference libraries based on the ORFs in bait and prey libraries (There are 542 bait and 316 prey ORFs in the case of this study).

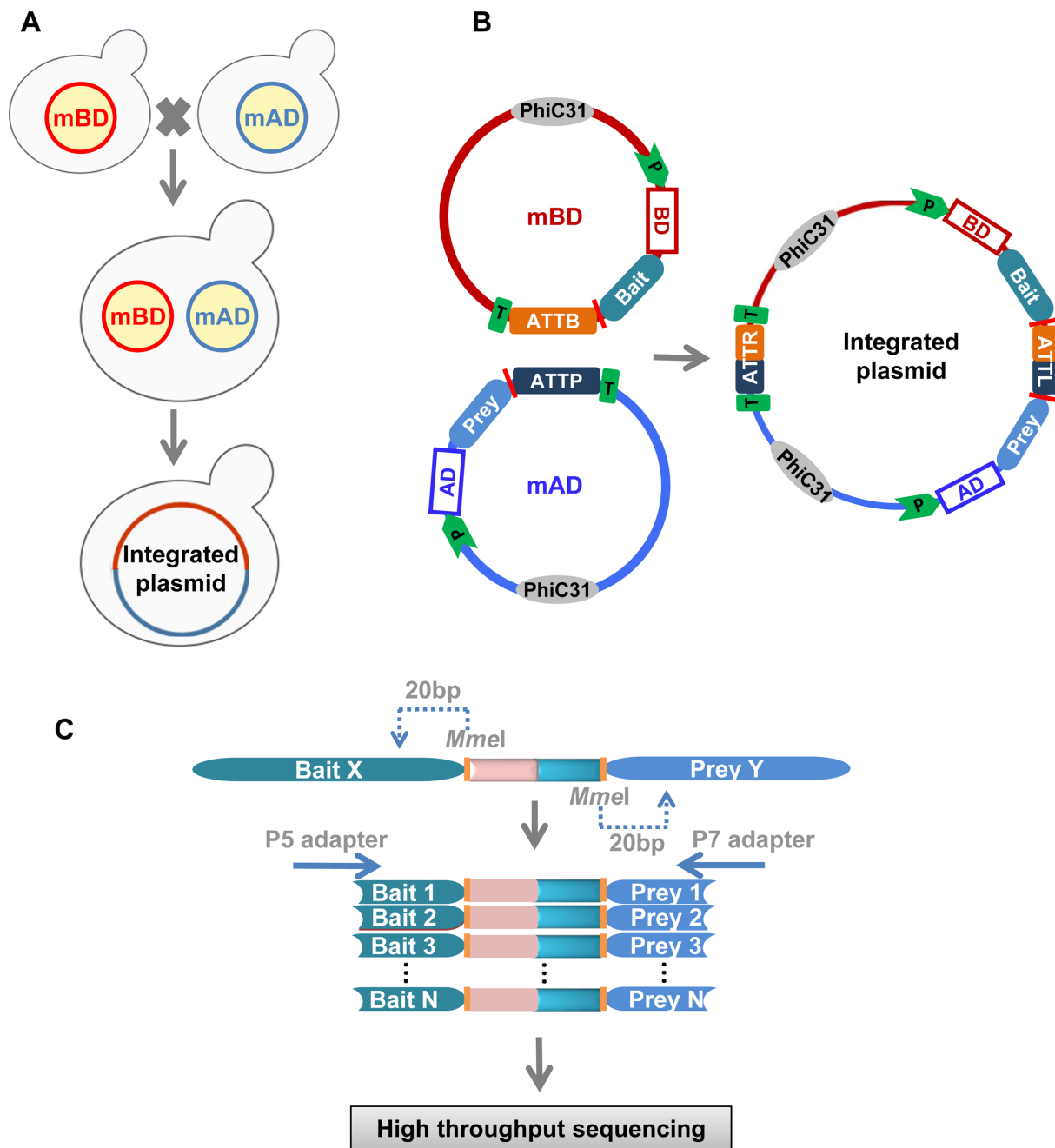
Align the trimmed sequencing pairs against the reference libraries using BLAST+. Only unique mapping reads with no mismatch and the sequence pairs in which both of reads could be aligned unambiguously with no mismatch are retained.

Interaction frequency of an interacting protein pair can be calculated by counting the sequence pairs that could be aligned to them.

## RESULTS

### Design of the technology

To establish a high-throughput system which allows simultaneous PPI screening between a bait library and a prey library, the bait and prey vectors of the conventional Y2H system from Clontech were modified to develop a novel recombination-based 'library versus library' Y2H system, designated RLL-Y2H (Figure 1). The design of this system is based on the high efficient DNA integration ability of PhiC31 as well as its specific recognition sites, ATTB and ATTP (22,23). Basically, ATTB and ATTP sites were fused right behind the bait and prey genes, respectively. In the presence of PhiC31 integrase, the bait and prey plasmids will be irreversibly integrated into



**Figure 1.** Flowchart of Recombination-based ‘Library vs Library’ Y2H system (RLL-Y2H) screening. (A) The yeasts with bait and prey plasmids are mated and only the yeasts containing interacting bait-prey pairs can grow on SD/-AHLT selection plates. The bait and prey plasmids are integrated into a single plasmid so that the interacting bait and prey genes are paired in a single DNA fragment. (B) The integrase PhiC31 recognition sites ATTB and ATTP are inserted right behind the bait and prey genes, respectively. PhiC31 is inserted into the backbone and is driven by the Pol III promoter. The integrase PhiC31 recognizes ATTB and ATTP sites and integrates the bait and prey plasmids into a big plasmid *in vivo* which unambiguously couples the interacting pair of specific bait and prey genes. The ADHI promoters and polyadenylation sites, as well as their positions after the recombination were indicated by the green bars with letter ‘P’ and ‘T’, respectively. The stop codons are indicated by the red lines right behind Bait and Prey. (C) The recombinated bait and prey fragments are amplified and digested by Type II restriction enzyme *MmeI* to generate ~110 bp fragments containing bait tail-ATTL-prey tail, followed by DNA sequencing library construction and high-throughput sequencing.

a single plasmid, in which the bait and prey genes are fused at tail-to-tail orientation with an ATTL linker in between (designated bait-ATTL-prey) (Figure 1A–C; Supplemental Figure S1A–C). In order to highly express PhiC31 and avoid competition with the ADH1 promoter which drives the bait and prey genes, we screened several *Saccharomyces cerevisiae* promoters and selected Pol III promoter to drive PhiC31 expression for its high activities. As illustrated in Figure 1B, PhiC31 together with its recognition sites ATTB and ATTP were inserted into pGBKT7 and pGADT7 vectors of the Matchmaker GAL4-based two-hybrid assay system to generate new bait and prey vectors named mBD and mAD. The derived plasmids containing bait and prey genes were introduced into diploid yeast cells via mating and selected on dropout plates lacking Adenine/Histidine/Leucine/Tryptophan (SD/-AHLT). The integration of bait and prey plasmids occurred in individual yeast cells *in vivo* and only the cells containing positive interacting protein pair could grow on the selective SD/-AHLT plates. The survived cells containing interacting bait-and-prey pairs were pooled for DNA extraction and PCR amplification. This approach avoids individual yeast isolation, DNA preparation and individual gene barcoding, therefore greatly reduces the laborious workload. Moreover, the sequences of the fused bait and prey amplicons can be used to unambiguously decode each interacting bait-and-prey pair and allow high throughput ‘library vs library’ PPI screening. The PCR amplicons of fused bait and prey pairs were mostly more than 1 kb with variable sizes, which are very challenging for high-throughput sequencing. To this end, type II restriction enzyme *MmeI* (which cuts DNA ~20 bp downstream of its recognition site) recognition sequences were inserted right after bait and prey genes, respectively, to generate a ~110 bp DNA fragments containing a pair of 20 bp interacting proteins cDNA strands and ATTL linker in between (Figure 1C). These fragments were then ligated with Illumina P5 and P7 adapters and subjected to high-throughput sequencing.

### Development of the technology

To test the efficiency of this RLL-Y2H system, positive controls consisting of a well-known interacting protein pair, murine p53 and SV40 T-antigen (24), were inserted into the mBD and mAD of RLL-Y2H system, respectively. The interaction was clearly detected (Figure 2A), and the transformation, mating efficiency as well as the interacting efficiency were compatible with the conventional Y2H system (data not shown). To investigate the integration efficiency, the survived colonies containing p53 bait plasmid and T-antigen prey plasmid on the SD/-AHLT selective plates were used to amplify the p53-ATTL-T-antigen integration fragment. As shown in Figure 2B, amplification of each colony yielded the expected PCR products, and the precise integration was further confirmed by Sanger sequencing (Figure 2C). Next, this p53-ATTL-T-antigen integration fragment was digested by *MmeI*, which generated an expected ~110bp fragment comprising of p53 *MmeI* end, ATTL linker and T-antigen *MmeI* end (Figure 2D). To-

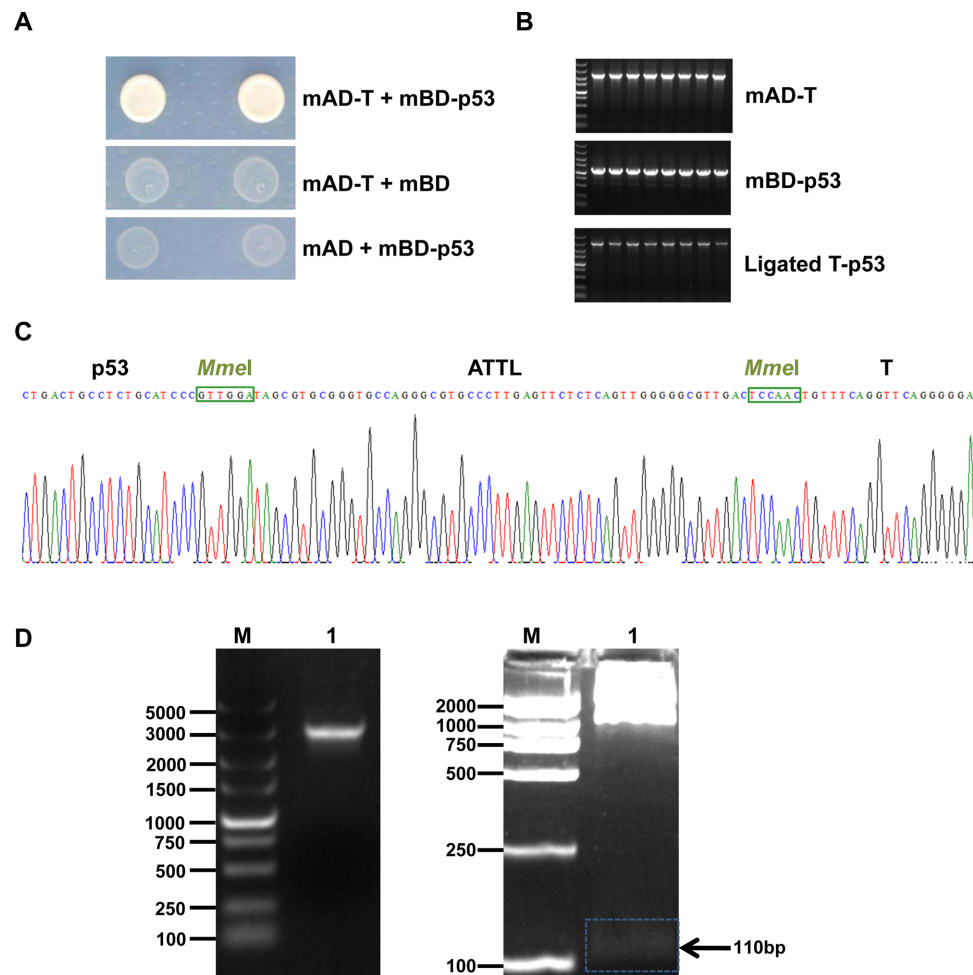
gether, these data provide proof of the concept for our novel high-throughput RLL-Y2H system.

### Construction and screening of *M.tb* bait library and host prey library

Next, we applied this RLL-Y2H system to investigate the interactome between *M.tb* and host cells, which occurs mostly between the membrane and secretory proteins of *M.tb* and immunity-related proteins, the autophagy pathway proteins and *M.tb* susceptibility genes of host cells. To this end, 542 *M.tb* (H37Rv) membrane and secretory protein ORFs (predicted by <http://tuberculist.epfl.ch> and <http://www.tbdb.org>) and 316 mouse immunity-related proteins, *M.tb* susceptibility genes, autophagy pathway protein ORFs (18–21) were cloned to construct the bait and prey libraries, respectively (Figure 3A, Supplemental Tables S1 and S2). Each bait and prey was tested for self-activation before screening. Twenty out of 542 baits were detected as self-activators, and were therefore excluded for the subsequent screenings (red marked in Supplemental Table S1). The bait library was mated with the prey library in 4 × 50 ml mating culture with ~17% mating efficiency, yielded ~8800 colonies on SD/-AHLT selective plates. All the colonies were pooled for plasmid extraction and PCR amplification. A smear band representing integrated bait-ATTL-prey fragments were observed (Figure 3B) and then harvested for *MmeI* digestion, which generated a ~110bp band comprising of the bait *MmeI* end, the ATTL linker and the prey *MmeI* end (Figure 3C). These digested products were ligated with Illumina P5 and P7 adapters to generate a high-throughput sequencing library (Figure 3D). To evaluate the reproducibility of the RLL-Y2H system, we performed two independent parallel screening from a small bait library containing 150 *M.tb* genes and a small prey library containing 200 host cell genes. To decode the protein interaction network between *M.tb* and host cell from the high-throughput sequencing data, a custom script was used to filter raw data, align reads and construct PPI network (Figure 3E). As shown in Figure 3F, 190 out of 227 of the paired interacting bait and prey sequences from screening 1# can be re-captured by an independent screening 2# (with 273 pairs).

### Delineation and confirmation of PPI network between *M.tb* and the host

Next, we performed the full screening with 542 ORFs of *M.tb* and 316 ORFs of host cells. Among 46 million raw reads, 14 million reads could be mapped as *M.tb*-host interacting protein pairs. We identified a total of 1,438 PPIs between 82 *M.tb* proteins and 128 host proteins (Figure 4A; Supplemental Table S3). The interaction frequencies of *M.tb*-host protein pairs were calculated based on the number of corresponding aligned reads. Of note, the number of the interaction pairs with low sequencing counts (<5) is dramatically higher than that of the others (Figure 4B). To obtain a high-quality protein interaction network, the interaction pairs with low counts (<5) were excluded from PPIs. Meanwhile, the distribution of host and TB proteins according to the number of their interactors showed that there



**Figure 2.** Proof-of-principle experiment of RLL-Y2H system. (A) Positive control interacting protein pair, murine p53 and SV40 T-antigen were inserted into mBD and mAD of RLL-Y2H system and were transformed into yeast respectively. Only the yeasts containing both p53 and SV40 T-antigen but not the controls can grow on SD/-AHL T selection plates. (B) After plasmid extraction from yeast, p53, SV40 T-antigen and their recombinated fragments were amplified by PCR. (C) Sanger sequencing of the recombinated p53 and SV40 T-antigen fragments. The linker ATTL and *MmeI* sites were highlighted. (D) The PCR products of the recombinated p53 and SV40 T-antigen fragments before (left panel) and after (right panel) *MmeI* digestion.

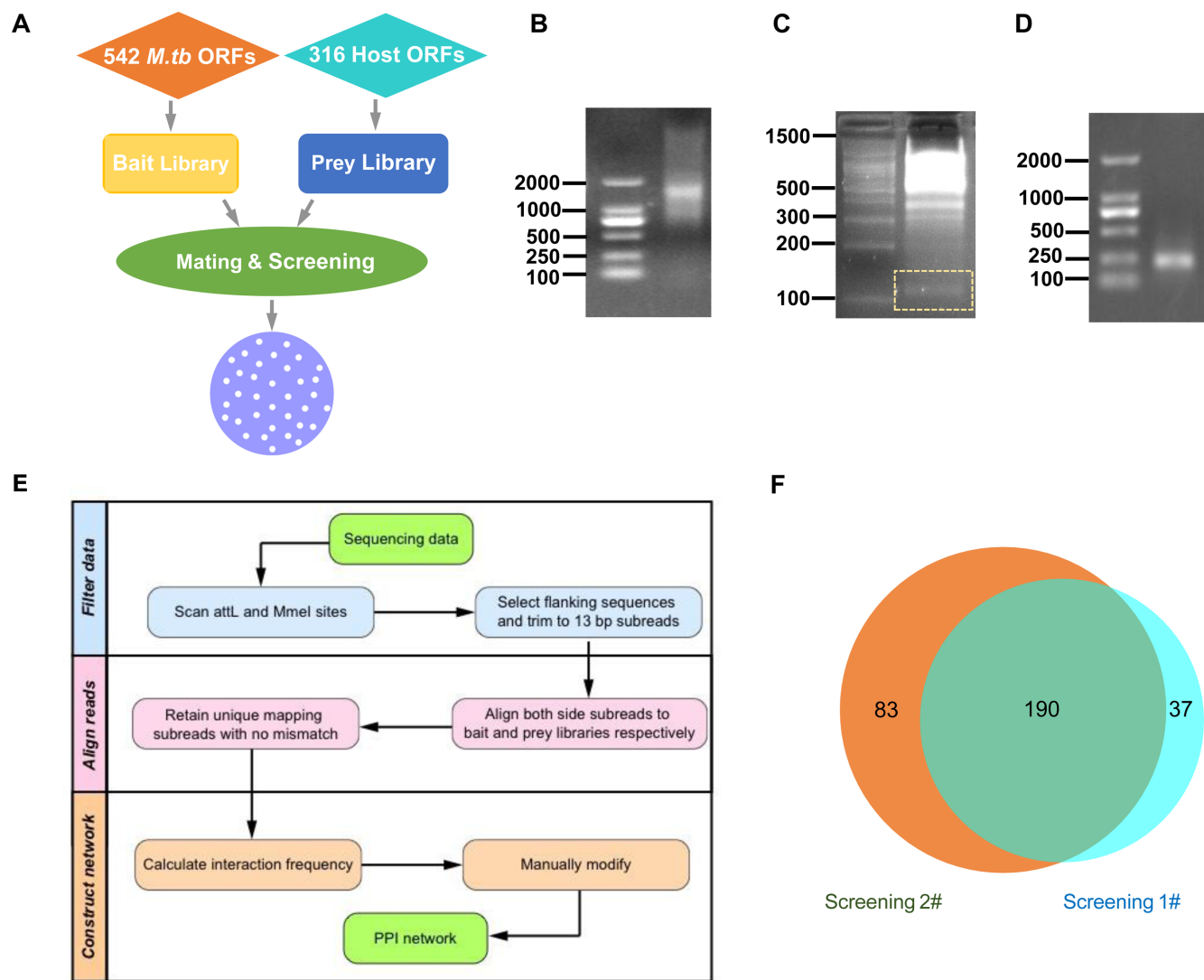
were long tails in both graphs indicating a small amount of proteins with a large number of interactors (more than 45, Figure 4C and D). Thus, the proteins with more than 45 interactors were also removed from the network, as they are most likely sticky proteins with high chance to physically but not physiologically interact with many proteins. Finally, we delineated a PPI network between *M.tb* proteins and host proteins containing 441 PPIs (Figure 4E; Supplemental Table S4). The host proteins were divided into several groups according to their physiological functions such as toll-like receptor signaling, T cell receptor signaling, Autophagy, *M.tb* susceptible genes, Jak/Stat signaling, NF- $\kappa$ B signaling and B cell receptor signaling. The PPI sub-networks belong to specific signaling pathways were then constructed to show their potential physiological function (Figure 5A and Supplemental Figures S2 and S3).

As each screening system can generate false positive data, the identified PPIs were then subjected to interaction retest. To systematically evaluate the false positive rate, 269 randomly selected protein interaction pairs were transformed into Y187 and Y2HGOLD yeast strains, respectively, and

were mated to confirm their interaction individually. As shown in Supplemental Figure S4 and Table S5, 189 out of 269 (70.3%) protein interaction pairs can be confirmed. To verify the cutoff value (five counts), we re-tested 40 interacting pairs with less than 5 counts. As shown in Supplemental Figure S5 and Table S6, only 6 out of 40 (15%) could be confirmed, which was much less than the average positive rate (70.3%). Meanwhile, the positive rate of re-test data from the randomly selected 269 interacting pairs were plotted against their sequencing counts. As shown in Supplemental Figure S5, the positive rate and the sequencing counts are positively correlated, suggesting that sequencing count 5 is a rational cutoff.

#### Identification of the *M.tb* proteins interfering autophagy in host cells

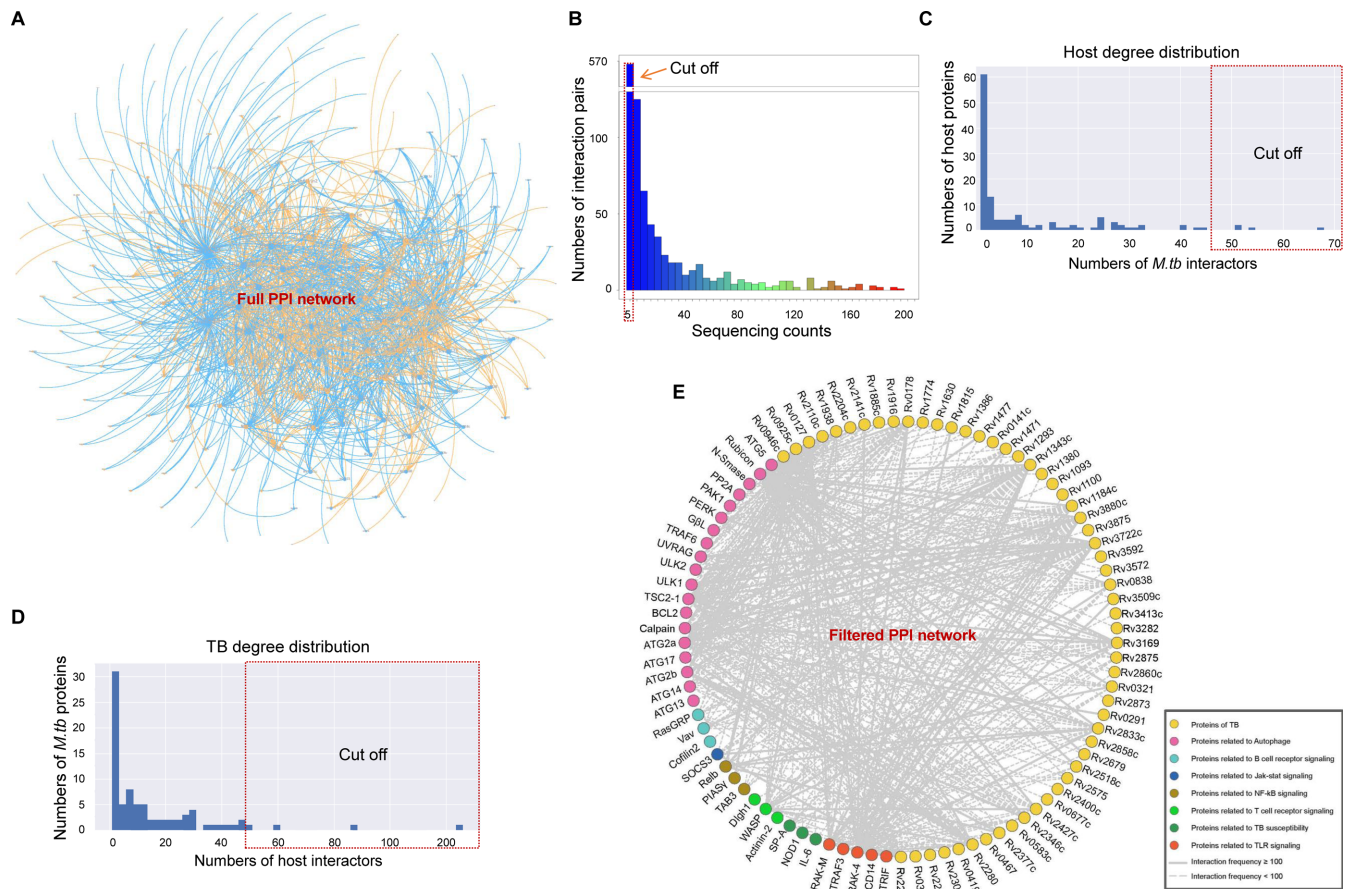
Next, we aimed to elucidate the potential physiological function of the PPI network we identified between *M.tb* and the host cells. As an intracellular pathogen, *M.tb* interferes with host cell defense machinery, such as autophagy path-



**Figure 3.** RLL-Y2H *M.tb* and host libraries construction and screening. (A) Flowchart of *M.tb* bait library and host prey library construction and screening. A total of 542 *M.tb* ORFs and 316 host protein ORFs were cloned and transformed into yeast Y2HGold and Y1871 strains, respectively. The bait library and prey library were mated for screening in a single pool. (B) After plasmid isolation from the positive colony pool, the recombined prey and bait fragments were amplified and (C) digested by *MmeI*. (D) Electrophoresis of the RLL-Y2H sequencing library. (E) Flowchart of sequencing data analysis and PPI network construction. (F) Venn diagram of the bait-prey pair library sequencing data from two independent screening.

way (25). Thus, the sub-set network between *M.tb* and autophagy pathway was extracted from the entire PPI network (Figure 5A). To test whether the *M.tb* proteins in this network could potentially interfere with host cell autophagy, we generated a stable cell line expressing GFP<sub>ph</sub>-mCherry-LC3 for autophagy assay. More specifically, eGFP were mutated to a pH super-sensitive version (26) and then fused with mCherry and LC3. This autophagy reporter was inserted into the H11site (27) of Raw264.7 cell genome using CRISPR/Cas system (Figure 5B; Supplemental Figure S6A and B). In this cell line, rapamycin could dramatically induce autophagy as indicated by the aggregation of GFP/mCherry dots and their decrease of GFP/mCherry fluorescence ratio which changed in response to the pH gradient due to phagosome-lysosome fusion (28). Next, 38 *M.tb* genes in this *M.tb*-autophagy PPI network and a

control (empty vector) were individually transfected into this stable cell line (Figure 5B). Autophagy was induced by rapamycin and then subjected to fluorescence assay. Of note, transfection of Rv2427c could significantly rescue the rapamycin-induced decrease of GFP/mCherry fluorescence ratio of aggregation dots, suggesting this protein can interfere with autolysosome formation (Figure 5C and D; Supplemental Figure S7). We also performed immunofluorescence experiment to confirm the expression of Rv2427c in GFP<sub>ph</sub>-mCherry-LC3 Raw cell line as shown in Supplemental Figure S7. According to our PPI network, four host autophagy pathway related proteins (PERK, Rubicon, TSC2 and WASP) have been confirmed to interact with Rv2427c. As the interaction frequency of PERK and TSC2 were dramatically lower than that of Rubicon and WASP, we further verified the interaction between Rv2427c and Ru-



**Figure 4.** Delineation the protein interaction network between *M.tb* and host cells. (A) The whole genome PPI network between *M.tb* bait and host cell. Blue nodes represent TB proteins and yellow nodes represent host proteins. (B) The distribution of TB–host protein interaction pairs according to their sequencing counts. 5 was chosen as cut-off as there was a dramatic increase of interaction pairs below 5 counts. (C and D) The distribution of host and TB proteins is according to the numbers of their interactors. The sticky proteins (>45 interactors) have higher degree than others and therefore were cut-off for future analysis. (E) The filtered PPI network between *M.tb* bait and host cell.

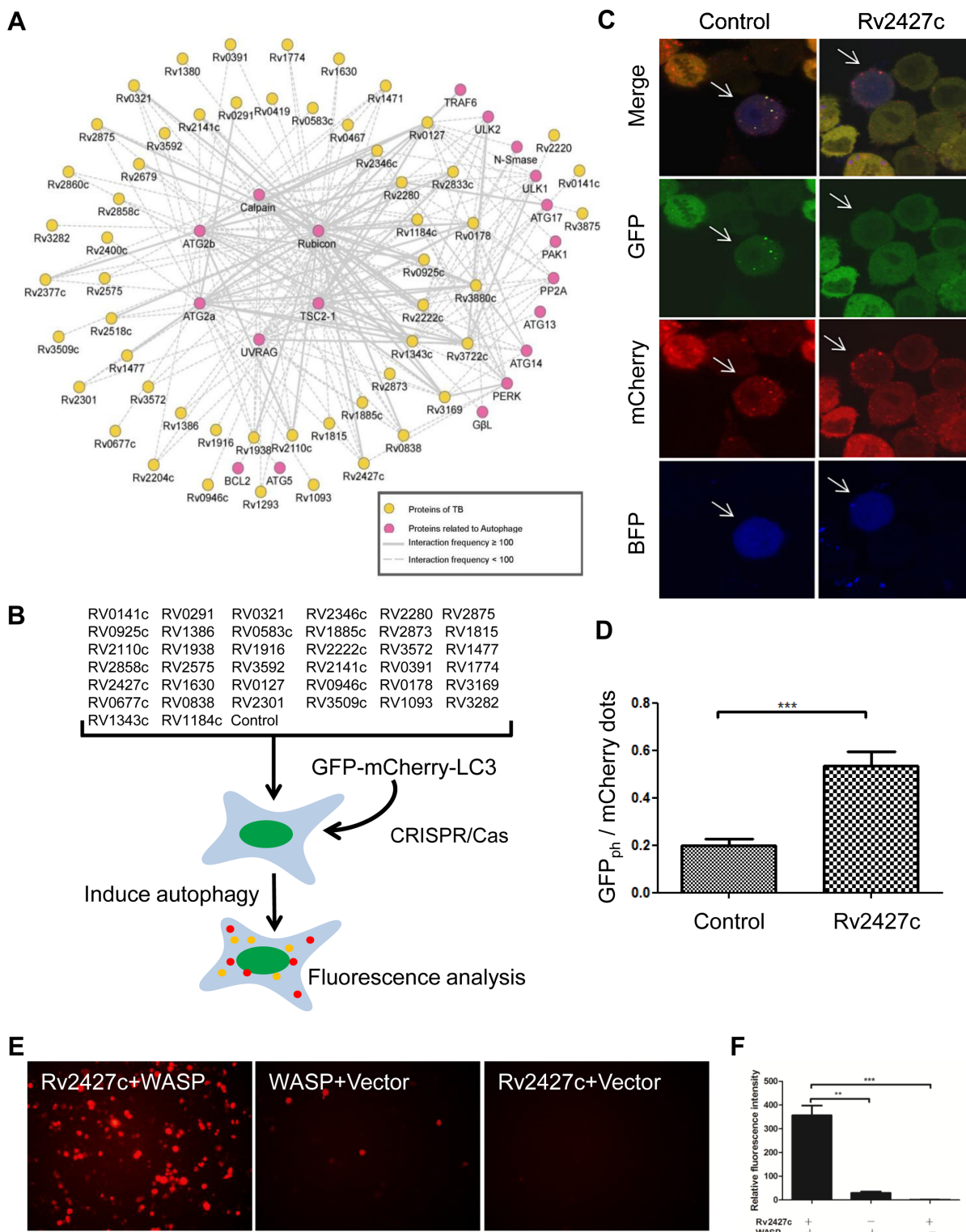
bicon and WASP by bimolecular fluorescence complementation (BiFC) assay in eukaryotic system. As shown in Figure 5E and F, the interaction between Rv2427c and WASP can be validated by the BiFC assay. However, the interaction between Rv2427c and Rubicon could not be confirmed by this approach and needs to be further investigated.

## DISCUSSION

Y2H is an ideal system to screen direct PPI but suffers a lot for its intensive labor. While recent-state-of-the-art high-throughput Y2H technologies, such as Smart-pooling Y2H, Y2H-Seq and Stitch-Seq Y2H systems reduced the Y2H screening efforts and cost (3,4,7–9). These systems require either individual arrays to prepare gene library, yeast transformation or individual treatment for plasmid extraction and PCR amplification which are still expensive and time-consuming. A recently reported BFG-Y2H system ingeniously tags each pair of bait and prey genes with a specific barcode and utilize Cre-lox system to fuse barcodes from distinct plasmids (9). As Cre-lox system mediated barcode fusion can be reversible, it may influence the efficiency of the system. Nevertheless, this system can easily decode the interaction network via reading these chimeric protein-pair

barcodes. However, it is still laborious and expensive when a large library with unique barcodes needs to be prepared. Recently, a CrY2H system was developed for all-by-all high-throughput PPI screening, in which the PCR products of the fused bait and prey genes were randomly sheared for sequencing (11). Although this system doesn't need barcoding individual gene, only 2.4% CrY2H reads contain sequence of both prey and bait genes. Thus, to delineate a big PPI network, a massive deep sequencing is required. In this study, the RLL-Y2H system took advantage of a highly efficient recombination activity of Phic31 integrase to efficiently and irreversibly combine a bait and a prey into a single plasmid *in vivo* (22,23). As the length of fused DNA fragments of interacting bait and prey genes are long and diverse, they are not suitable for high-throughput sequencing. To this end, we inserted *MmeI* restriction sites on both ends of bait and prey genes, which could release a 110bp fragment after digestion. This fragment is suitable for high-throughput sequencing, of which the sequences containing matrix information of interacting bait and prey pair can be directly used to decode PPI networks without any barcoding procedure. Together, our simple RLL-Y2H system could substantially reduce screening time, cost and labor work, and has the-





**Figure 5.** Identification of the *M.tb* proteins interfering autophagy in host cells. **(A)** The PPI network between *M.tb* and autophagy pathway proteins of host cells. **(B)** Raw264.7 cells expressing GFP<sub>ph</sub>-mCherry-LC3 for autophagy assay was generated using CRISPR/Cas9 system. The stable cell line transfected either with an empty vector or with 38 *M.tb* genes in the *M.tb*-autophagy PPI network individually, were treated with rapamycin to induce autophagy. **(C)** Rv2427c and control plasmid were transfected to GFP<sub>ph</sub>-mCherry-LC3 Raw264.7 stable cell line and induce autophagy by rapamycin. Transfection of Rv2427c significantly rescued the rapamycin-induced decrease of GFP/mCherry fluorescence ratio of the aggregation dots. Arrows indicate transfection positive cells. Arrows indicate transfection positive cells. **(D)** Quantification of GFP/mCherry fluorescence ratio of the aggregation dots (*n* = 48). **(E)** Verification of the protein-protein interaction by BiFC assay. Rv2427c were co-transfected with WASP and the negative controls into HEK 293T cells, respectively. Cells were harvested 36h after transfection for confocal fluorescence microscopy-based image analysis. The interactions between Rv2427c and prey proteins allow re-formation of a bimolecular fluorescent complex. **(F)** Quantification of the fluorescence.

oretically the capacity for genome-wide massive library vs library screening. Compared with the approaches mentioned above, our system has several advantages: 1) No requirement for individual treatment such as plasmid isolation and PCR amplification; 2) No requirement for gene barcoding which allows big library vs library screening at once for huge PPI network; 3) The application of *MmeI* digestion can efficiently remove the massive deep sequencing noise. As an intracellular pathogen, *M. tb* has equipped with various machinery to evade host immune system in order to replicate and persist within host cells. A trans-kingdom interactome is hence evolved when *M. tb* proteins encounter host proteins. The center of this interactome is the interaction network between *M. tb* secretory and membrane proteins and the host immunity- and autophagy-related proteins. Here, we applied the RLL-Y2H system to decipher the central PPI network between *M. tb* and host cells and to investigate the physiological role of the subset interaction network between *M. tb* and host autophagy pathway. Interestingly, Rv2427c was identified to interfere with phagosome-lysosome fusion, which may play an important role in perturbing host autophagy pathway. It has been shown that WASP is crucial for lysosome motility (29), and can directly bind to ezrin which is required for the efficient fusion of lysosome and phagosome (30). Therefore, it could be possible that Rv2427c may interfere the physiological function of WASP by direct protein-protein interaction and thus inhibit autolysosome formation. It would be of great importance to further investigate the physiological role of these *M. tb* proteins in counteracting autophagy pathway, such as initiation, elongation, maturation and recycling. These will shed a light on the pathogenesis of *M. tb* and may facilitate the identification of anti-*M. tb* drug targets.

### Limitation

One general shortcoming of high throughput screening is the high false positive rate. In our study, 70.3% of PPI were confirmed by independent retransformation of the 269 randomly chosen pairs of the identified interacting proteins. In two previous large-scale human interactome screening studies using traditional Y2H system, the authors could verify 65% and 78% of the identified interactions, respectively (6). A recent study developing CrY2H-seq system for PPI screening demonstrated a positive rate of 73% (11). Our positive rate is comparable with previous studies, suggesting that the ~29.7% false positive rate is unlikely due to the modification of the conventional Y2H system or interference by the PhiC31. It is rather a common feature of Y2H screening which usually yields a certain amount of false positive interacting proteins (31–33). In this scenario, we showed that 190 out of 227 pairs of the interacting protein pairs can be re-captured by independent parallel screening. It would be ideal if the PPI screening system could also be applied to quantify the protein interaction strength. Theoretically, this can be semi-quantified by the frequency of sequencing reads, as the strong interactors generally grow faster than those weak interactors. However, it is challenging at this stage, mostly due to the fact that the amplification of the integrated plasmids could be biased. Since the length of the fused bait and prey fragments before *MmeI* digestion

is very diverse, ranging from ~600 bp to 5 kb. Compared to the long fragments of the fused bait-prey, the short fragments might be most likely over-amplified due to the bias of PCR amplification. This problem might be solved by using emulsion PCR which could independently amplify the individual DNA on beads (34). It is thus possible to update RLL-Y2H as a semi-quantitative screening system to map the PPI strength by the frequency of sequencing reads. Every technology has its own most suitable application. In this regards, RLL-Y2H is more suitable for large-scale library vs library interaction screening. For medium scale PPI network screening, array-based Y2H such as eY2H, Mini-pool Y2H or Smart-pooling Y2H might be a better option. In our system, although the gene coding region will not be affected after recombination, the lack of polyadenylation sites may affect the expression of bait and prey proteins. Nevertheless, in our positive control experiments using P53-T pair, the yeast could grow properly on selective plates, suggesting that the expression of bait and prey genes were not significantly affected. This could be attributed to several reasons: (i) the bait, prey and fusion plasmids are all present simultaneously in a zygote yeast; (ii) the fusion of bait and prey plasmids happens after the expression of phic31, during this time the bait and prey genes are independently expressed and can thus initiate the expression of the selective marker genes if they bind to each other; (iii) even there are no polyadenylation sites after the bait and prey genes in fusion plasmids, these genes can still be expressed at a certain level. In future studies, this system could be updated by moving the polyadenylation sites to the fronts of ATTB and ATTP sites (Supplemental Figure S8), so that the expression of all bait and prey genes would not be affected at all after recombination. In our study, we plotted the numbers of the host or TB proteins against the number of their interactors and found long tails in both graphs. These proteins located in the end of tails are most likely sticky proteins with high chance to physically but not physiologically interact with many proteins, and were therefore removed from PPI network. It would be ideal to experimentally verify this cutoff value. However, while we can design a rational experiment to verify the lowest cutoff value, there is no perfect experiment to validate the highest cutoff value. The cutoff value needs to be carefully selected for different experiments depending on the sequencing depth, overall positive rate and some other factors.

### Perspective

Other than Y2H for PPI screening, this recombination-based 'library vs library' screening concept can also be applied to other yeast screening systems such as Y1H (35,36) and Y3H (37,38). These approaches may open new avenues for delineating the global PPI, protein–DNA and protein–RNA interaction landscape and might pave the way to understand the details of the signaling processes in cells.

### AVAILABILITY

The accession number for the sequencing data reported in this paper is GEO: GSE93036.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors are grateful for Dr Ow for providing PhiC31 plasmid. We would like to thank N. Zhang for FACS experiments and Tyagi for proof-reading and discussion.

*Authors contributions:* G.C and F.Y conceived the RLL-Y2H idea. F.Y, Y.L, M.Z, Q.Y, X.W, C.Z, R.T, X.C and D.W performed experiments. Y.H and W.X performed the bioinformatics analysis. S.X and H.C provide some experiment materials. G.C., F.Y. and Y.L. wrote the manuscript, with inputs from all other authors. All the authors discussed the results and commented on the manuscript.

## FUNDING

National Key Research and Development Program of China [2017YFD0500303]; National Natural Science Foundation of China [31371106, 91640105]; Doctoral Fund of Ministry of Education of China [131012]; China Agriculture Research System [CARS-36]; Huazhong Agricultural University Scientific & Technological Self-innovation Foundation [52204-13002]. Funding for open access charge: National Key Research and Development Program of China [2017YFD0500303]; National Natural Science Foundation of China [31371106, 91640105].

*Conflict of interest statement.* None declared.

## REFERENCES

- Yates, J.R. (2000) Mass spectrometry. From genomics to proteomics. *Trends Genet.*, **16**, 5–8.
- Rajagopala, S.V. and Uetz, P. (2009) Analysis of protein-protein interactions using array-based yeast two-hybrid screens. *Methods Mol. Biol.*, **548**, 223–245.
- Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., Svrzikapa, N., Hirozane-Kishikawa, T., Rietman, E., Yang, X. *et al.* (2011) Next-generation sequencing to generate interactome datasets. *Nat. Methods*, **8**, 478–480.
- Rolland, T., Tasan, M., Charletoaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R. *et al.* (2014) A proteome-scale map of the human interactome network. *Cell*, **159**, 1212–1226.
- Lopez, J. and Mukhtar, M.S. (2017) Mapping protein-protein interaction using high-throughput yeast 2-hybrid. *Methods Mol. Biol.*, **1610**, 217–230.
- Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Beriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N. *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173–1178.
- Xin, X., Rual, J.F., Hirozane-Kishikawa, T., Hill, D.E., Vidal, M., Boone, C. and Thierry-Mieg, N. (2009) Shifted Transversal Design smart-pooling for high coverage interactome mapping. *Genome Res.*, **19**, 1262–1269.
- Weimann, M., Grossmann, A., Woodsmith, J., Ozkan, Z., Birth, P., Meierhofer, D., Benlasfer, N., Valovka, T., Timmermann, B., Wanker, E.E. *et al.* (2013) A Y2H-seq approach defines the human protein methyltransferase interactome. *Nat. Methods*, **10**, 339–342.
- Yachie, N., Petsalaki, E., Mellor, J.C., Weile, J., Jacob, Y., Verby, M., Ozturk, S.B., Li, S., Cote, A.G., Mosca, R. *et al.* (2016) Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.*, **12**, 863.
- Hastie, A.R. and Pruitt, S.C. (2007) Yeast two-hybrid interaction partner screening through in vivo Cre-mediated Binary Interaction Tag generation. *Nucleic Acids Res.*, **35**, e141.
- Trigg, S.A., Garza, R.M., Macwilliams, A., Nery, J.R., Bartlett, A., Castanon, R., Goubil, A., Feeney, J., O'Malley, R. and Huang, S.C. (2017) CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nat. Methods*, **14**, 819.
- WHO (2016) *Global Tuberculosis Report 2016*. 2016WHO Press.
- Via, L.E., Deretic, D., Ulmer, R.J., Hibler, N.S., Huber, L.A. and Deretic, V. (1997) Arrest of mycobacterial phagosome maturation is caused by a block in vesicle fusion between stages controlled by rab5 and rab7. *J. Biol. Chem.*, **272**, 13326–13331.
- Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeon, L., Abdallah, A.M., Caleechurn, L., Bottai, D., van Zon, M., de Punder, K. *et al.* (2012) ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cell Microbiol.*, **14**, 1287–1298.
- Romagnoli, A., Etna, M.P., Giacomini, E., Pardini, M., Remoli, M.E., Corazzari, M., Falasca, L., Goletti, D., Gafa, V., Simeone, R. *et al.* (2012) ESX-1 dependent impairment of autophagic flux by *Mycobacterium tuberculosis* in human dendritic cells. *Autophagy*, **8**, 1357–1370.
- Walburger, A., Koul, A., Ferrari, G., Nguyen, L., Prescianotto-Baschong, C., Huygen, K., Klebl, B., Thompson, C., Bacher, G. and Pieters, J. (2004) Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science*, **304**, 1800–1804.
- Abràmoff, M.D., Magalhães, P.J. and Ram, S.J. (2004) Image processing with ImageJ. *Biophoton. Int.*, **11**, 36–42.
- Bellamy, R., Ruwende, C., Corrah, T., McAdam, K.P., Whittle, H.C. and Hill, A.V. (1998) Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *N. Engl. J. Med.*, **338**, 640–644.
- Rossouw, M., Nel, H.J., Cooke, G.S., van Helden, P.D. and Hoal, E.G. (2003) Association between tuberculosis and a polymorphic NFκB binding site in the interferon γ gene. *Lancet*, **361**, 1871–1872.
- Ferwerda, G., Girardin, S.E., Kullberg, B., Le Bourhis, L., De Jong, D.J., Langenberg, D.M., Van Crevel, R., Adema, G.J., Ottenhoff, T.H. and Van der Meer, J.W. (2005) NOD2 and toll-like receptors are nonredundant recognition systems of *Mycobacterium tuberculosis*. *PLoS Pathog.*, **1**, e34.
- Pan, H., Yan, B., Rojas, M., Shebzukhov, Y.V., Zhou, H., Kobzik, L., Higgins, D.E., Daly, M.J., Bloom, B.R. and Kramnik, I. (2005) Ipr1 gene mediates innate immunity to tuberculosis. *Nature*, **434**, 767–772.
- Kudriavtseva, E.A., Yun, Y.L., Voeikova, T.A. and Lomovskaia, N.D. (1994) Use of a plasmid with integrative function of phage phiC31 for transfer of cloned genes into *Streptomyces* strains. *Genetika*, **30**, 886–897.
- Thorpe, H.M. and Smith, M.C. (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 5505–5510.
- Li, B. and Fields, S. (1993) Identification of mutations in p53 that affect its binding to SV40 large T antigen by using the yeast two-hybrid system. *FASEB J.*, **7**, 957–963.
- Duan, L., Yi, M., Chen, J., Li, S. and Chen, W. (2016) *Mycobacterium tuberculosis* EIS gene inhibits macrophage autophagy through up-regulation of IL-10 by increasing the acetylation of histone H3. *Biochem. Biophys. Res. Commun.*, **473**, 1229–1234.
- Mahon, M.J. (2011) pHluorin2: an enhanced, ratiometric, pH-sensitive green fluorescent protein. *Adv. Biosci. Biotechnol.*, **2**, 132–137.
- Tasic, B., Hippenmeyer, S., Wang, C., Gamboa, M., Zong, H., Chen-Tsai, Y. and Luo, L. (2011) Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 7902–7907.
- Kimura, S., Noda, T. and Yoshimori, T. (2007) Dissection of the autophagosomal maturation process by a novel reporter protein, tandem fluorescent-tagged LC3. *Autophagy*, **3**, 452–460.
- Chang, F.S., Stefan, C.J. and Blumer, K.J. (2003) A WASp homolog powers actin polymerization-dependent motility of endosomes in vivo. *Curr. Biol.*, **13**, 455.
- Marion, S., Hoffmann, E., Holzer, D., Clainche, C.L., Martin, M., Sachse, M., Ganeva, I., Mangeat, P. and Griffiths, G. (2011) Ezrin promotes actin assembly at the phagosome membrane and regulates phago-lysosomal fusion. *Traffic*, **12**, 421–437.
- Caufield, J.H., Sakhawalkar, N. and Uetz, P. (2012) A comparison and optimization of yeast two-hybrid systems. *Methods*, **58**, 317–324.

32. Kuchaiev,O., Rašajski,M., Higham,D.J. and Pržulj,N. (2009) Geometric de-noising of protein-protein interaction networks. *PLoS Comput. Biol.*, **5**, e1000454.
33. Vidalain,P., Boxem,M., Ge,H., Li,S. and Vidal,M. (2004) Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods*, **32**, 363–370.
34. Margulies,M., Egholm,M., Altman,W.E., Attiya,S., Bader,J.S., Bemben,L.A., Berka,J., Braverman,M.S., Chen,Y.J., Chen,Z. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.
35. Gstaiger,M., Knoepfel,L., Georgiev,O., Schaffner,W. and Hovens,C.M. (1995) A B-cell coactivator of octamer-binding transcription factors. *Nature*, **373**, 360–362.
36. Reece-Hoyes,J.S. and Walhout,A.J.M. (2012) Gene-centered yeast one-hybrid assays. *Methods Mol. Biol.*, **812**, 189–208.
37. SenGupta,D.J., Zhang,B., Kraemer,B., Pochart,P., Fields,S. and Wickens,M. (1996) A three-hybrid system to detect RNA-protein interactions in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8496–8501.
38. Putz,U., Skehel,P. and Kuhl,D. (1996) A tri-hybrid system for the analysis and detection of RNA–protein interactions. *Nucleic Acids Res.*, **24**, 4838–4840.